

Dataset Documentation

Marine benthic invertebrate Dataset (Biosys)

17/06/2022

This document will help you understand and use the Marine benthic invertebrate dataset held in Biosys

Dataset description

This is for Approval for Access product Afa130.

BioSys is the Environment Agency database for storing, manipulating and reporting data from freshwater and marine biological surveys at any taxonomic level. This monitoring work is undertaken by the Environment Agency and by third parties.

All the relevant data associated with Marine benthic invertebrate site, sample and analysis information can be obtained from the [Ecology and Fish data explorer](#). The data can be accessed using the interactive map to produce small downloadable files or they can be downloaded in bulk as a set of relational CSVs that can be linked via unique codes in each file. Note: this dataset may be referred to as 'TraC' (Transitional and Coastal water) or 'marine' benthic invertebrates but the terms mean the same thing for this dataset.

The data download is delivered in one CSV file within the Ecology and Fish data explorer. You will see it labelled in the download as follows:

- BENTH_OPEN_DATA_TAXA (Contains benthic invertebrate site, sample and taxa information)

If you are bulk downloading the full dataset there is also a second file available which contains additional taxonomic information for all of the taxa listed in our database:

- OPEN_DATA_TAXON_INFO (Contains taxonomic information for all taxon names available in Biosys)

The tables can be related to each other using the joins below:

- OPEN_DATA_TAXON_INFO.TAXON_LIST_ITEM_KEY = BENTH_OPEN_DATA_TAXA.TAXON_LIST_ITEM_KEY

Please refer to the dataset content table, at the end of this document, for an explanation of the fields/columns in each file.

API

The benthic invertebrate dataset can also be accessed using an API (Application Programming Interface): [Ecology and Fish Data API](#). For more information regarding the service please visit the dedicated Frequently Asked Questions section of the Defra Data Services Portal (DSP) Forum: [Ecology and Fish Data FAQs](#).

This dataset includes data from monitoring locations in England only. For Welsh survey data please contact [Natural Resources Wales](#).

Update frequency

This dataset is extracted monthly on the first Sunday of each month. It can then take a week or so for the data to appear on the explorer. Data entered onto Biosys immediately prior to extraction may not have been quality assured and so may not reflect the final version of the data. Recent data can be identified by viewing the DATE_OF_ANALYSIS field.

Common questions & known issues

Known issues:

Reporting Area and Agency Area names:

The Reporting Area field was added to the Explorer as part of the 2022 update. The Reporting Area identifies the team who report on that data. It is more closely aligned to the [Environment Agency public face Area names](#) than the older, Agency Area field. The table below shows the comparison between all the fields.

Please be aware that the geographical boundaries for the old Southern and Thames Agency Areas do not match the geographical boundaries for the current Areas replacing them. Therefore, a small proportion of sites located in the old Area may not be located in the matched new Area:

Agency Area (Old EA Region and Area Name)	Reporting Area	EA Public Face Area Name
Anglian - Central	East Anglia - West	East Anglia
Anglian - Eastern	East Anglia - East	East Anglia
Anglian - Northern	Lincolnshire and Northamptonshire	Lincolnshire and Northamptonshire
North West - North	Cumbria and Lancashire	Cumbria and Lancashire
North West - South	Greater Manchester Merseyside and Cheshire	Greater Manchester Merseyside and Cheshire
South West - Devon	Devon and Cornwall	Devon and Cornwall
South West - Cornwall	Devon and Cornwall	Devon and Cornwall
South West - North Wessex	Wessex	Wessex
South West - South Wessex	Wessex	Wessex
Midlands - Central	West Midlands	West Midlands
Midlands - East	East Midlands	East Midlands
Midlands - West	West Midlands	West Midlands
North East - Yorkshire	Yorkshire	Yorkshire
North East - North East	North East	North East
Thames - North East	Hertfordshire and North London	Hertfordshire and North London
Thames - South East	Kent South London and East Sussex	Kent South London and East Sussex

customer service line
03708 506 506

incident hotline
0800 80 70 60

floodline
0345 988 1188

www.gov.uk/environment-agency

Thames - West	Thames	Thames
Southern - Kent & E. Sussex	Kent South London and East Sussex	Kent South London and East Sussex
Southern - Solent & South Downs	Solent and South Downs	Solent and South Downs

Some marine sites will be listed against '**National Marine**' for the Agency Area. That does not correspond to any specific 'Area' in the country. Some 'Reporting_Area' fields may be listed as 'unknown' if the site is located a significant distance offshore.

WFD Waterbody ID:

Care should be taken when using this field to search for data. The WFD Waterbody ID field is populated on Biosys with the Water Framework Directive waterbody ID that the site *sits* in. This may be different to the Waterbody that the site is used to produce a WFD status assessment for. This field is often null for historic sites and may also be null where a site sits in a waterbody that is too small to have a WFD ID. The ID may also relate to the WFD cycle 1 set of waterbodies or cycle 2 waterbody IDs. If you are interested in a particular waterbody ID then it is also worth using this field in conjunction with a search on the interface map and/or an extract of the data using the Waterbody name.

The [catchment data explorer](#) can also be used to give further high level information on a particular waterbody and could be used to compliment/support interpretation of this dataset.

Common questions:

What data are included in this dataset?

All marine benthic invertebrate samples that have been collected and analysed in the laboratory are included in this dataset.

Please note that we do not include pollution sample data or protected species where releasing details of their location could have a detrimental effect e.g., freshwater pearl mussels.

Site name is excluded from this dataset. This is because it is a free-text field and may contain information or data that cannot be released under an Open Government Licence.

The sample reason field in this dataset displays a generic primary sample reason such as 'National monitoring' or 'Local investigative monitoring'. It does not include detail on all of the reasons for which the sample has been collected.

How are samples collected?

Subtidal sampling: Survey vessels deploy grabs and corers, taking sediment of a known area to collect quantitative, subtidal benthic samples from coastal and estuarine waters. The Environment Agency (and our predecessor's) standard sampling device for marine and estuarine work is a 0.1 m² day grab, which is reliable and simple to use.

Intertidal sampling: In some circumstances it is possible to obtain quantitative intertidal benthic samples using a grab from a boat at high-tide. If this is not possible, sites are instead accessed at low tide on foot or by using a RIB (Rigid-hulled Inflatable Boat) or a hovercraft. Samples are then taken using cores or box cores. Historically, we (and our predecessors) have predominantly used small hand corers (0.01 m²), employing large box cores (0.1 m²) on some occasions, particularly on more sandy sediments.

3-Minute pond net sample: For a small proportion of our samples the method used to collect the samples will be a 3-Minute kick sample with a pond net. This is usually a freshwater sampling technique but may

have been used in some estuarine waterbodies where it is possible to access the site and collect a sample using this method. This method is not comparable with the other methods listed above.

Why do you collect replicate samples?

Due to the heterogeneity of marine and estuarine communities, more than one sample will be required in a defined habitat type to provide adequate information to describe the communities. So multi-replicate or multiple, random-stratified sampling is necessary. Each individual sample will have a different sample_id and replicate code.

How are the samples analysed?

Marine benthic invertebrate samples are sent to external contractors for analysis. Each taxon will have an abundance recorded against it. If the taxon was present but no abundance was recorded this field will be null. Taxon qualifiers are also recorded, where required, to give additional information about the taxon e.g., the size or life stage.

Analysis of benthic invertebrate samples usually requires further supporting parameters such as Particle Size Analysis (PSA) and/or sediment chemistry. Where collected this data is archived in the [Water Quality Archive](#). The BENT_PSA field should denote if a PSA sample has been collected. The WIMS_SITE_ID will display the corresponding WIMS sample point.

Why are there both site and sample grid references?

In Biosys we store multiple samples against a site. For many of our marine survey techniques we record an NGR at the site because this acts as a central/representative location around which we collect lots of samples. An NGR is then also recorded against each of the samples to give a more accurate sampling location. In the map interface of the Ecology and Fish data explorer, if a sample has a sample NGR then that NGR is used to display the point on the map, otherwise the default is to use the site NGR.

How do I use the taxonomic information?

The Natural History Museum maintain the [UK Species Inventory \(UKSI\)](#) and we use that to update our 'taxon dictionary' in Biosys. They maintain a bespoke checklist for us that limits the taxon names we can record against and also ensures we have the taxon hierarchy and names required by our calculations. This does mean that sometimes we have bespoke names or names that are slightly out-of-date. You can use our dataset to understand the latest name that we use (as this may be different in older samples), the type of taxon (e.g. macrophyte) but also the recommended NBN Taxon_Version_Key. If you wish to link to other datasets then it is the taxon_version_key (TVK) that you need.

Disclaimers:

Note: We do our best to avoid quality problems but this dataset reflects the data we hold. Our datasets may contain errors.

External websites:

We have provided links to external websites as a convenience and to provide additional information. We are not responsible for the reliability or content of those external websites. We are not liable for any loss or damage that may come from using the external website links in this document.

Sort code:

We list the Sort code in Biosys which is derived from the Maitland list, owned and maintained by [CEH](#). Our sort code list contains omissions, additions and errors and should *not* be used as a definitive list. It can be used as an aid to ordering taxa in a taxonomic hierarchy.

Dataset content

Description of column headings

Field name	Description	Marine benthic invertebrate taxa	Taxon information
AGENCY_AREA	EA Area name as given in Biosys for site (e.g. NORTH WEST - NORTH)	Yes	
REPORTING_AREA	EA Area name based on more-up-to-date operational Areas e.g. Cumbria and Lancashire.	Yes	
SEA AREA	Sea Area in Biosys, free-text field (e.g. LIVERPOOL BAY)	Yes	
WATERBODY_TYPE_DESCRIPTION	Waterbody type recorded in Biosys from a pick-list (e.g. ESTUARY: Lower section of a river between freshwater limit and the sea)	Yes	
WATER_BODY	Water body in Biosys, free-text field (e.g. RIBBLE OFFSHORE)	Yes	
SITE_ID	Unique numeric Site identifier generated when the site was created in Biosys (e.g. 148190).	Yes	
SITE_VERSION	Version number of site in Biosys. Together with the SITE_ID this denotes a unique site record.	Yes	



SITE_NGR_PREFIX	Two letter prefix of National Grid Reference (e.g. SS) for the site.	Yes	
SITE_EASTING	Five-digit Easting of NGR (e.g. 58450) for the site.	Yes	
SITE_NORTHING	Five-digit Northing of NGR (e.g. 03430) for the site.	Yes	
SITE_NGR_10_FIG	Ten-digit National Grid Reference (e.g. SS5845003430) for the site.	Yes	
SITE_FULL_EASTING	Full Easting of the site (e.g. 258450) for the site.	Yes	
SITE_FULL_NORTHING	Full Northing of the site (e.g. 103430) for the site.	Yes	
WIMS_SITE_ID	This is the site reference for our WIMS water quality database e.g. LC544405. Data from WIMS can be found on the Water Quality Archive pages	Yes	
WFD_WATERBODY_ID	WFD ID code given in Biosys (e.g. GB112072065980).	Yes	
SAMPLE_ID	Sample ID generated when the sample is added to Biosys (e.g. 567642)	Yes	
SAMPLE_VERSION	Version number of sample in Biosys. Together with the SAMPLE_ID this denotes a unique sample record.	Yes	
REPLICATE_CODE	A free-text field used to differentiate between replicate samples taken at the same site, date and time.	Yes	

customer service line
03708 506 506

incident hotline
0800 80 70 60

floodline
0345 988 1188

www.gov.uk/environment-agency

SURVEY_CODE	Unique survey code that links a subset of surveys.	Yes	
SAMPLE_NGR_PREFIX	Two letter prefix of National Grid Reference (e.g. SS) for the sample NGR	Yes	
SAMPLE_EASTING	Five-digit Easting of sample NGR (e.g. 58450)	Yes	
SAMPLE_NORTHING	Five-digit Northing of sample NGR (e.g. 03430)	Yes	
SAMPLE_NGR_10_FIG	Ten-digit National Grid Reference of the sample location (e.g. SS5845003430)	Yes	
SAMPLE_FULL_EASTING	Full Easting of the sample location (e.g. 258450)	Yes	
SAMPLE_FULL_NORTHING	Full Northing of the sample location (e.g. 103430)	Yes	
SAMPLE_DATE	Date the sample was taken (e.g. 05/05/1994)	Yes	
SAMPLE_TYPE_DESCRIPTION	The type of sample collected (e.g. MARINE: PHYTOPLANKTON - Truncated List)	Yes	
SAMPLE_METHOD_DESCRIPTION	The type of sample method used (e.g. MARINE: Surface water sample)	Yes	
SAMPLE_REASON	Displays a generic sample reason e.g. National/Investigative/Local purposes etc.	Yes	
BENT_GRAB_DEPTH	The depth that a sediment grab has been taken at on a benthic invertebrate survey	Yes	

customer service line
03708 506 506

incident hotline
0800 80 70 60

floodline
0345 988 1188

www.gov.uk/environment-agency

BENT_PSA	A yes/no option to denote if a particle size analysis sample has been taken in conjunction with the benthic invertebrate sample (PSA data relating to sediment grain size is stored on WIMS).	Yes	
BENT_DEPTH_RPD_LAYER	The depth of the Redox Potential Depth (RPD) layer (CM)	Yes	
ANALYSIS_ID	Analysis ID created when analysis has been added against the sample (e.g. 562376)	Yes	
DATE_OF_ANALYSIS	Date the sample was analysed.	Yes	
ANALYSIS_TYPE_DESCRIPTION	Analysis type description as recorded in Biosys e.g. LABORATORY PRIMARY: Analysed in laboratory by primary analyst	Yes	
ANALYSIS_METHOD_DESCRIPTION	Analysis method description as recorded in Biosys e.g. SEDIMENTATION CHAMBER: Sedimentation chamber (algae counts)	Yes	
SIEVE_SIZE	The sieve mesh size used in the analysis of the sample.	Yes	
TAXON_LIST_ITEM_KEY	This is a unique ID, assigned by the NHM, to each name in the EA checklist. This code is required by Biosys.	Yes	
TAXON_NAME	Name of the taxon unit e.g. <i>Tubificoides benedii</i>		Yes

customer service line
03708 506 506

incident hotline
0800 80 70 60

floodline
0345 988 1188

www.gov.uk/environment-agency

TAXON_QUALIFIER_DESC	Additional information describing an attribute of the taxon such as the size or life stage e.g. Juvenile	Yes	
NUMBER_FOUND	Number of individuals found	Yes	
IS_THIRD_PARTY_DATA	Lists whether the data is from a third-party organisation. Note: a 'No' entry may still mean data has been provided by a third-party (particularly for historic data) as this is a relatively new field in Biosys.	Yes	
TAXON_VERSION_KEY	This is a unique ID, assigned by the NHM to different versions of a taxon name e.g. where authority or context (e.g. <i>sensu stricto</i>) is different. This is the NBN code that can be used to link to data with other organisations & the National Biodiversity Network.		Yes
AUTHORITY	Naming authority e.g. (L.) Gaertn.; Linnaeus, 1758.		Yes
SORT_CODE	This is a code (also known as the maitland code) defined by CEH to enable us to sort taxa in a hierarchy. NOTE: our list is not the definitive list and contains omissions/errors. The latest list can be downloaded from the CEH website.		Yes
TAXON_NAME_CURRENT	Yes/No field to denote where the taxon is the most current name in use on our system.		Yes

customer service line
03708 506 506

www.gov.uk/environment-agency

incident hotline
0800 80 70 60

floodline
0345 988 1188

PREFERRED_TAXON_NAME	This is the most current and preferred name for the taxon on Biosys. If the TAXON_NAME_CURRENT field is 'No' then you will see the latest name listed here.		Yes
PREFERRED_NAME_TVK	The Taxon Version Key for the Biosys preferred taxon name. This is the NBN code that can be used to link to data with other organisations & the National Biodiversity Network.		Yes
NBN_RECOMMENDED_TVK	This is the NHM recommended Taxon Version Key (taken from their Nameserver table). This is the NBN code that can be used to link to data with other organisations & the National Biodiversity Network.		Yes
PARENT_TLIK	The parent Taxon_List_Item_Key for the parent taxon in Biosys		Yes
PARENT_TAXON_NAME	The Parent taxon name.		Yes
TAXON_RANK	Taxon rank e.g. Family, species etc.		Yes
TAXON_TYPE	Lists the taxon type in Biosys e.g. Algae; Other Macroinvertebrate; Macrophyte.		Yes
TAXON_GROUP_NAME	Lists the taxon group as defined by the NHM e.g. alga; crustacean.		Yes
NON_NATIVE_SP	Lists whether the taxon is a non-native species to England.		Yes

customer service line
03708 506 506

www.gov.uk/environment-agency

incident hotline
0800 80 70 60

floodline
0345 988 1188

PROTECTED_TAXA	Lists whether the taxon is a protected species. Note we use JNCC information to help us list protected species but we only flag taxa that have the potential to be impacted by activities we permit or activities we carry out.		Yes
-----------------------	---	--	-----

customer service line
03708 506 506

www.gov.uk/environment-agency

incident hotline
0800 80 70 60

floodline
0345 988 1188